DOCUMENT RESUME

ED 188 189 CS 205 614

AUTHOR Southwell, Michael G.
TITLE Evaluating Writing Courses and Programs: Some Practical Management and Design Considerations.
PUB DATE Mar 80
NOTE 11p.; Paper presented at the Annual Meeting of the Conference on College Composition and Communication (31st, Washington, DC, March 13-15, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Course Evaluation; *Evaluation Criteria; Higher Education; *Program Evaluation; Student Attitudes; *Student Evaluation; Teacher Attitudes; *Writing (Composition); *Writing Instruction
IDENTIFIERS Writing Apprehension

ABSTRACT

A model for designing a writing program evaluation is based on the assumptions that the entire writing course or program is evaluated, not just student writing; that the evaluation is summative rather than formative; and that the role of the evaluator is clearly defined. The procedure for implementing the evaluation should include the following steps: (1) pilot-testing the evaluation instrument; (2) evaluating student writing using holistic scoring or descriptive measures; (3) evaluating student attitudes using standard writing apprehension tests, questionnaires, interviews, or participant-observer studies; (4) evaluating teacher attitudes using questionnaires or interviews supplemented by student evaluations and administrative observations; (5) evaluating administrator attitudes using interviews; (6) analyzing the transferability of the course (the ability of the course to be taught successfully by many different teachers); (7) considering demographic information such as pass and retention rates; and (8) implementing a longitudinal study of the effects of instruction. (AEA)

EVALUATING WRITING COURSES AND PROGRAMS: SOME PRACTICAL MANAGEMENT AND DESIGN CONSIDERATIONS

CCCC/1980, Washington, D.C.
14 March 1980

Michael G. Southwell
Department of English
York College, CUNY
Jamaica, NY 11451

In beginning to plan an evaluation, you must answer three important questions.

1. What do you want to evaluate? One common answer is student writings. Various kinds of measures for evaluating writings are available, and I'll discuss some of them later. But another possible answer (and a better one) is the entire course (or program) in which students produced those writings. An evaluation of a course or a program would certainly include evaluation of writings, but would go far beyond it.

It's important to remember that a program that produces successful writing can't necessarily be called successful itself. For example, a course that greatly improves student writing but at the same time causes very high drop-out rates could hardly be called successful. Or what about a course that is highly teacher-dependent? A course that can be taught by only one teacher, or by only one certain kind of teacher, doesn't seem successful as a model course. I would encourage you, then, to consider going beyond simple evaluation of student writings; much

more can be learned if you evaluate an entire course. The rest of my discussion will focus chiefly on this kind of program evaluation.

2. Why and for whom are you evaluating it? If you are interested in finding out more about how an existing program is working, in order to improve it, then you are engaging in what is called formative evaluation. You yourself are the audience for such an evaluation.

On the other hand, you may be interested in an evaluation for others (like administrators, or a funding agency, or a local or regional education agency). This kind of evaluation (called summative evaluation) usually is used to provide information on which a decision about whether or not to continue a program (often a new or innovative one) can be based. For such an evaluation, it will not be enough just to describe the program; it should be compared to other existing programs. It will very likely be necessary to set up a comparison-group design for this kind of evaluation, using experimental and control groups, and pre- and post-test measures. Such an evaluation is complex and expensive, but can provide a great deal of information about how successful a program is. In the rest of my discussion, I'll assume that you are a program developer who is planning such a summative evaluation of a writing course.

3. Who will be evaluating it? You yourself (as a program developer) are likely to have the deepest understanding of the issues involved, but it might be difficult for you to be completely objective about the results (and it will certainly be difficult for others to believe that you can be objective). Having outsiders do the evaluation will assure more objectivity, but will also raise the possibility of an inappropriate evaluation (as I will discuss shortly).

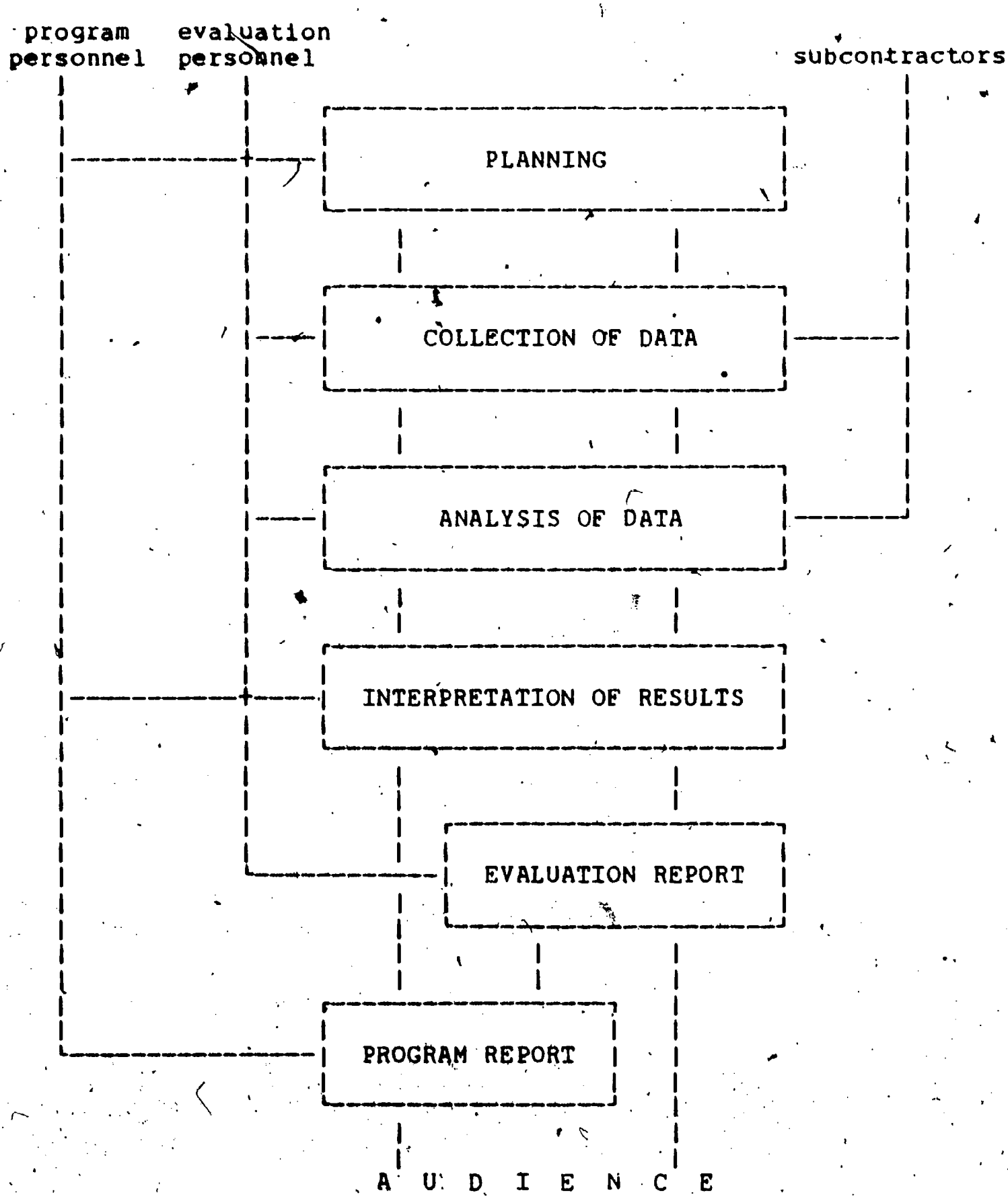
Once you have answered these questions, you are ready to begin planning for the evaluation itself. Some words about the philosophy of such evaluation are in order here. Evaluation of educational programs is never the clear-cut scientific evaluation portrayed in statistics texts. The variables are simply too many and too complex and too uncontrollable. The performance of students in a course is probably affected as much by politics, or sports schedules, or the weather, or events in their (or their teachers') families, as it is by instruction. It is certainly possible that the variables most affecting the students' performance can't even be identified. This means that it will be very difficult to reach any definite conclusions about why different programs produce different results. Because scientific verification of theories is so unlikely, the emphasis should be on the discovery of facts. If you can get a full and complete description of what actually happened, it will be possible at least to speculate about why it happened. At the same time,

there will be an implicit suggestion that the reconstruction of similar-enough conditions could produce similar results; but educational systems are so complex that you should not expect to attain scientific proof (which implies repeatability).

If it is decided that a summative evaluation of a program is to be carried out by an outsider, it's important for you as a program developer to protect your program, both against an inappropriate evaluation (that is, measuring the wrong things, or measuring the right things wrongly), and against inappropriate conclusions from the evaluation. The management model shown in Figure 1 can provide this kind of protection.

In this model, the program developers and the evaluators collaborate on planning the evaluation. The evaluators collect the data, and analyze it (using subcontractors if necessary). The program personnel collaborate in interpretation of the results. The evaluators then write their evaluation report, and the program developers write a separate program report, comprehending the evaluation report. This model makes it possible for program personnel (who know most about what a program is supposed to do) to provide input at the stages where it is necessary, but at the same time preserves the objectivity of the data collection and analysis. Separate reports provide both program developers and evaluators with an opportunity to present their own interpretations of the results.

Figure 1: A Management Model for Program Evaluation



I can't emphasize too strongly that the first step in planning

any evaluation like this absolutely must be the specification of these details of the collaboration in a formal written contract. The preliminary discussion essential for establishing such an agreement with the evaluators will clarify for both of you just what needs to be done, and will also avoid potential problems by assigning responsibility for carrying out the activities of the evaluation.

I turn now to a discussion of what measures might be appropriate to evaluation of an entire writing program. It is important that a variety of measures be employed, for this will improve the validity of the evaluation greatly (by increasing the chances of picking up any changes that might have occurred).

To the extent possible, all evaluation measures planned should be pilot-tested, either by you or by the evaluators. It is astonishing how many unanticipated problems can arise in an evaluation; pilot-testing makes it possible to forestall at least some of them. It's useful to think of pilot-testing as a kind of formative evaluation not of the instruction but of its evaluation. This suggests that a trial run of the proposed evaluation during a previous semester can serve two purposes: as a pilot-test of the proposed evaluation, and as formative evaluation of the program. This is the most efficient way to pilot-test.

1. Evaluation of student writings is, of course, the most important single measure. Descriptive measures (such as an error count, a word count, or a t-unit count) might be used, if appropriate. Holistic or primary-trait scoring can provide a measure of the overall quality of student writings. A forced-choice scoring offers an easy way to compare writings, but is too crude a measure to be very valuable by itself. A combination of different measures will probably be able to provide more useful results than any single measure could. A good source for discussion of the various kinds of measures available is Charles Cooper and Lee Odell, Evaluating Writing (NCTE, 1977).

2. Evaluation of student attitudes is especially important when a new instructional method is being evaluated. It is important to investigate both students' attitudes toward their instruction, and their attitudes toward themselves as writers. You should try to avoid students' becoming aware that they are in a special or experimental program, for this is likely to affect their performance as well as their attitudes.

Pre- and post-administrations of a standardized writing apprehension test might be useful.

Questionnaires are the simplest way to get information, but need to be used carefully. The language should be checked for clarity and appropriateness. It might be difficult to obtain

sufficient numbers of responses for the results to be very informative. Different questions on the same subject offer a way to check the internal reliability of the responses, but there is always the problem of to what extent one can trust students' expressions of their feelings. Correlation of students' attitude responses with more objective measures (like their writing evaluation scores) might provide an index of how reliable their responses are.

Interviews can provide invaluable open-ended information, but require skillful and objective interviewers (students must not be threatened by interviews with anyone involved in their instruction). It may be difficult to get suitable interviewees, with accurate and representative perceptions about the program. Extreme care needs to be used in attempting to generalize about the information gained from interviews.

Important information can be gained by observing students at work, but it is likely to be even more valuable to use a participant-observation study. In this, a trained evaluator both observes students at work and participates in their work with them. A participant-observer can provide sophisticated information about both the work and the students' experiences in doing that work.

3. Teacher attitudes should also be measured, particularly if a reorganization of pedagogy is involved. Questionnaires and interviews can be supplemented by student evaluations and administrative observations or evaluations (if these exist).

4. The attitudes of adminstrators should also be considered. It is important to be aware of what institutional demands (in terms of costs, scheduling, facilities, etc.) a program makes. This is information that can probably be gained only by means of interviews.

5. An analysis of the transferability of the course is also often important. Can it be taught successfully by many different teachers? Is a large amount of training necessary? Comparing attitudes of subpopulations of teachers is one good way to gain this kind of information; it may also be possible to break down other evaluation results among subpopulations (comparing the writing scores of one group of students to those of another, for example).

6. Demographic information should not be overlooked. The traditional demographic categories (age, gender, previous college history, standardized test scores, etc.) may provide a clue to differences in results. Pass rates and retention rates are not usually thought of as demographic information, but they are among the most important pieces of information to have. They should be

used, however, only in the context of rates established over several semesters; otherwise, there is the danger of drawing seriously misleading conclusions. For example, low pass rates may result not from an inferior program, but from a certain teacher's well-established grading tendencies; or spring semesters may always have lower pass rates than fall semesters; etc.

7. Finally, a <u>longitudinal</u> study of the effects of the instruction over time should be carried out, if possible. Although such a study takes lots of time and money, it is the only way to determine whether a writing course has any effect which is more than just temporary.

I hope that some of these suggestions will be able to help you to make sure that evaluations of your own courses and programs give you information that is both accurate and helpful.